Fermilab

UPC No. 140

Survey of the Ring, a Trilogy;

Part 1: MAGNET POSITIONS, MATRIX PARTITIONS, AND MOORE-PENROSE

Leo Michelotti

September 23, 1980

Abstract

We propose using the Moore-Penrose pseudoinverse and singular value decomposition to process ES/D magnet survey data. For the case of a maximally symmetric circular ring these become equivalent to performing harmonic analysis on the data, but they have the advantage of being applicable to arbitrarily configured lattices. Using the pseudoinverse, we examine the effects of random measurement errors on closed orbits. This suggests a criterion for deciding whether survey data are sufficiently good to justify repositioning magnets.

Section 1 Introduction

A careful survey of the Fermilab main ring has been completed recently, and it is now incumbent on us to do something constructive with these data. The markings for the survey comprised a set of small, circular disks, what Thornton Murphy calls "sticky-backs", that were glued to the main ring floor in (presumably) well defined positions relative to the old station marks, many of the latter being no longer visible due to the installation of ES/D superconducting magnets. These new sticky-backs are to serve as sign posts for the Tevatron ring; accordingly, it is of some importance to know where they are.

The answer to that question lies buried in the survey data. Bringing it to the surface is complicated by the observation that the measurements are all relative: specifically, sagittal offsets and next-nearest-neighbor chords. Without a referent external to the ring it is impossible to calculate "absolute" positions, and this must be taken into account. Traditionally, this has been done by Fourier analyzing the data relative to site number around the ring and either throwing away or tailoring information in the low harmonic terms. The power of applying Fourier analysis to this problem, or to most other problems, derives from its ability to decouple a large set of equations. Essentially, each harmonic of the measurement data is excited by the

same harmonic of positioning errors and no other. Thus, a single complicated problem is divided into a large set of easy subproblems that can be solved in closed form. However, the drawback of this approach is that it is strictly correct only when the magnets are designed to be symmetrically positioned on a circle. It is not strictly correct for a more general configuration: a single Fourier harmonic in the measurement data will in general influence all positional harmonics, and vice versa.

Fortunately, Fourier analysis possesses natural extensions that are more general. One is embodied here in finding the Moore-Penrose pseudoinverse of the measurement Jacobian, or geodetic matrix: $\partial(\text{data})/\partial(\text{positions})$. In the next two sections of this memo we will formulate the problem more precisely, define the pseudoinverse, and briefly discuss some of its useful properties.

Random errors in the survey data will induce residual errors in the final placement of magnets which, unlike the original errors, will be correlated from site to site. These in turn will affect the closed-beam orbit, and it is of some interest to examine the relationship between orbit deviations and measurement errors. This is done in Sec.4 for a simple model in order to illustrate the method. Use of the pseudoinverse in this problem was tested numerically by applying it to the symmetric, circular ring; the results are given in Sec.5.

Now, a word about numbers that appear in the left-hand margin. My first impulse upon proofreading a memo that has just been typed almost invariably is to tear it up and start over. There are reasons that make this impractical, not the least of which is the transient nature of this document: It simply is not worth rewriting. Nonetheless, there are scattered here and there various instances of inadequate explanations, misleading or ambiguous phraseology, dual usage of a symbol, and so forth. Therefore, as a blanket covering for such gaffes, a number of retrospective comments are collected together in Sec.6 and cross indexed by numbers in the left-hand margin of the memo. Think of them as extended footnotes.

Section 2 The geodetic matrix

We will make a distinction between the magnets' "positions" and their "sites", the latter referring to their designed positions. These are labelled P_k, k = 1...N with indices increasing counter clockwise around the ring. Because the ring is periodic, it will be convenient later to assume the index set obeys a mod N arithmetic, so that $P_{k+N} = P_k$. It is not necessary that the sites be symmetrically located, nor even that they lie on a circle. For the present discussion, it is convenient for them to be at the vertices of a convex polygon, but even this is not indispensable.

"Ring" refers to the set of all magnets being surveyed - even if this is only a subset of all the magnets in the

accelerator - and its "state" to their positions. To keep matters simple, we shall confine our deliberations to the horizontal plane, but the formalism can be generalized easily to three dimensions.

The ring's global state is to be estimated by patching together data from a large number of local relative measurements. There is some freedom in selecting the local geometric quantitities to be used as survey variables. Usually in these problems the sagittal offset, D_k, and next-nearest neighbor chord length, a_k, are measured at each site P_k, as illustrated in Fig.(1). These could be supplemented or replaced by alternates such as nearest-neighbor chords, x_k, or angles between the magnets' lines of sight. It is not likely that the latter would be competitive with D_k measurement because of the large base lines in the triangles. (For example, nearest-neighbor chords in the main ring are $\sim$98' in length. One would have to measure θ_k to about 2 sec of arc in order for it to compete with a direct measurement of D_k made with $\sim$10 mils accuracy.) Measuring nearest-neighbor chords would be of some use in damping out high azimuthal harmonics, as will be seen in Sec.5. For the present, however, we shall consider the set $\{(D_k, a_k)\}$ to contain the primary survey variables.

The accuracy with which the ring's state can be estimated is very sensitive to <u>how</u> one uses the information in the data set. In the absence of measurement errors, these 2N

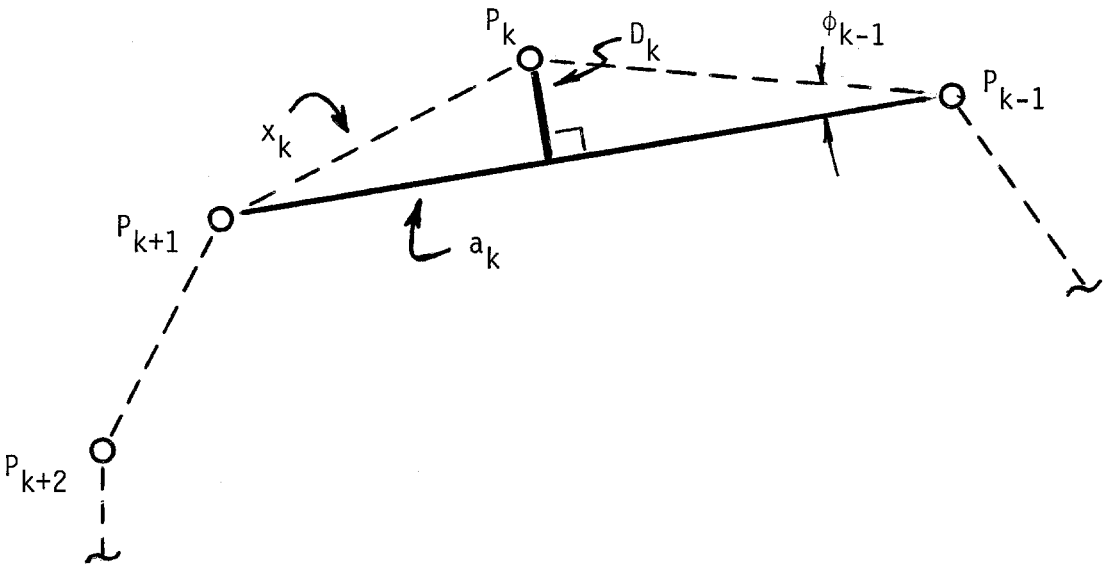

Fig. 1. Primary survey variables. Only part of a ring of sites is illustrated. Sagittal offset D_k and next-nearest neighbor chord a_k are indexed to correspond with site P_k. Nearest neighbor chords, x_k, and angles, ϕ_k, may be measured also, but these are treated as secondary variables, acting as redundancy checks on the primary ones.

quantities would "almost always" suffice to locate the positions exactly to within an isometry, i.e. rigid translations and rotations of the entire ring. (For the special case of an even number of sites symmetrically placed on a circle, a single nearest neighbor chord x_k would also be needed; we will see this explicitly later.) For example, one could reconstruct the ring by recursively determining P_{k+1} from P_k, P_{k-1}, and the data (D_k, a_k). The first position P_0 and the direction from P_0 to P_1 must be assigned arbitrarily, and it would be necessary to measure x_0 directly in order to place P_1 correctly. The recursion would then go around the ring, dropping off positions in its wake, until it returns to P_0 where it closes on itself and stops. Unfortunately, this method is unstable in that small measurement errors results in large deviations in positions even for moderate value of N, say $N \geq 30$. The ring would not close.

(1) To avoid this problem, we will solve for all positions simultaneously rather than sequentially. This is not as hard as it seems, because there already exists a design solution that can be used as a zeroth order approximation. Expanding to first order about the design linearizes the problem, making it computationally tractable. Essentially, differential displacements of the magnets result in differential deviations of survey measurements from their ideal values, those that they would have if the magnets were positioned at their sites. To use this approach

practically we must assume that the errors in originally placing the magnets are small so that the linear approximation is reasonable.

A convenient way to coordinatize the problem is to place a local (u,v) frame F_k at each site, oriented so that one axis, say the v axis, is aligned along $P_{k-1}P_{k+1}$. The other, the u axis, is orthogonal to it and points outward. The positions of all the magnets is recorded in the state vector

$$w = (u_1, v_1; u_2, v_2; u_3, v_3; \ldots; u_N, v_N)^T.$$

The state w is an element of a real vector space, the "state space" W of the ring; that is W obeys the usual closure axioms of a vector space: the sum and difference of states are states, and multiplying a state by a real number produces another legitimate state.

At this point it is useful to introduce a block notation that will be convenient later. We will use a subscripted semicolon to separate "outer" from "inner" indices: outer indices will refer to the sites; inner, to the coordinates. The 2N dimensional state vector w is indexed accordingly,

$$w_{1;k} = u_k \qquad \text{and} \qquad w_{2;k} = v_k \quad .$$

If an index (usually the inner) is suppressed, then the

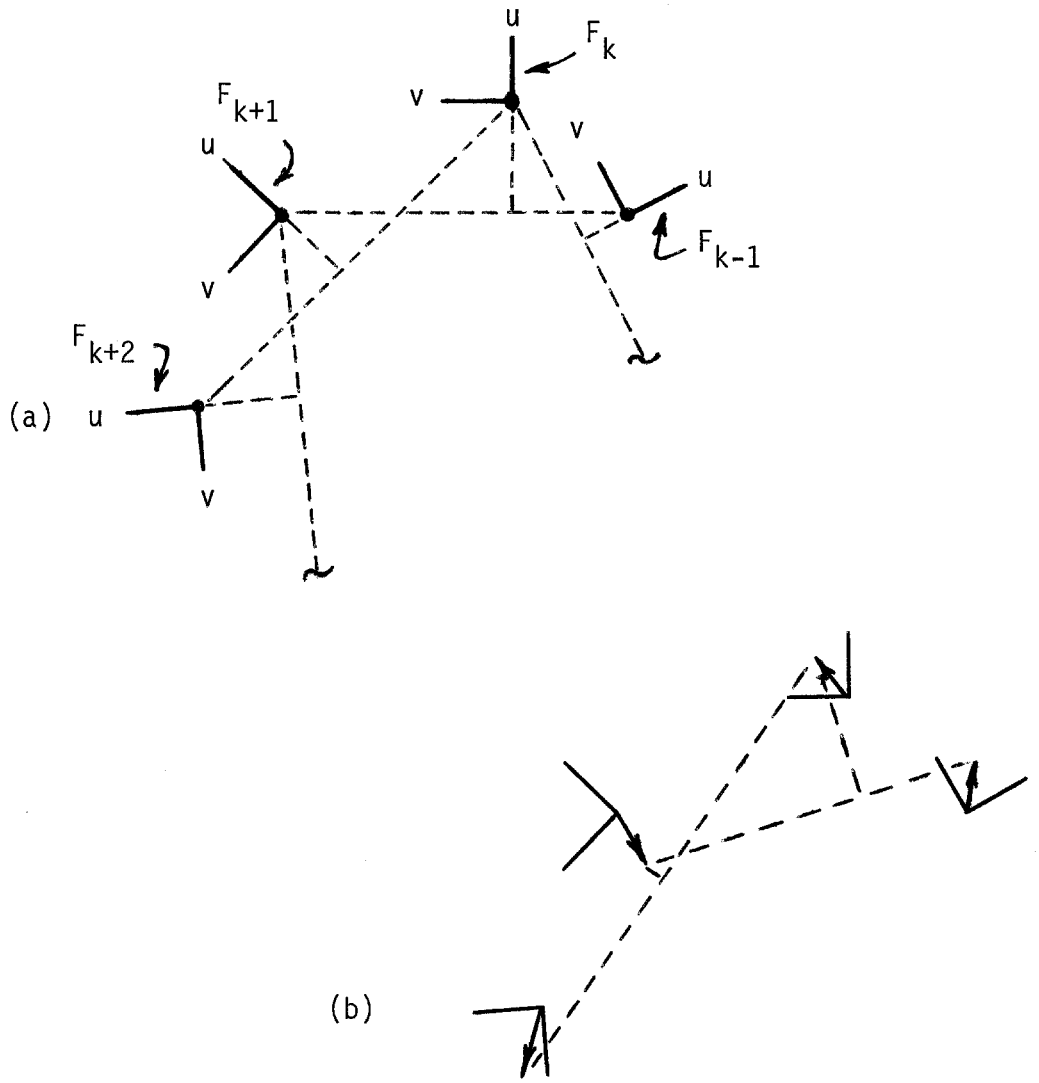

Fig. 2. Frames and the state of the ring. (a) A frame, F_k, is placed at each site in such a way that its axes are aligned along the directions of the primary variables. (b) Positions of the survey's markers (magnets, sticky-backs, or whatever) are coordinatized relative to these frames. Because the markers are not on their sites, due to placement errors, the measured values of the survey variables will be slightly different from their designed values.

corresponding sub-vector or array is understood to be represented. Suppression is accomplished either by blanking or insertion of a dot • into the index's location. Thus,

$$w_{;k} = w_{\bullet;k} = \begin{pmatrix} u_k \\ v_k \end{pmatrix},$$

$$w_{\bullet;\bullet} = w = \underline{\text{col}}\ (w_{;1}\ w_{;2}\ w_{;3}\ \cdots\ w_{;N}).$$

(Here "col" merely serves as a reminder that w is a 2Nx1 column matrix, although it is written as a row matrix to save space - a motivation that has just been frustrated by its explanation.) This notation will be helpful in dealing with block matrices.

$$A_{ij;k\ell} \qquad i,j = 1\ldots2, \qquad k,\ell = 1\ldots N.$$

According to the above conventions, $A_{;k\ell}$ would represent a 2x2 matrix, $A_{\bullet j;k\ell}$ a 2x1 array, and $A_{i\bullet;k\ell}$ a 1x2 array.

Associated with each state w of the ring is a set of measurement data,

$$M \equiv \underline{\text{col}}\ (M_{;1}\ M_{;2}\ M_{;3}\ \cdots\ M_{;N}) \in R^{2N}$$

$$M_{;k} = \underline{\text{col}}\ (D_k\ a_k) \in R^2.$$

Although it looks formally like a vector, M does not belong

to a vector space: sums and differences of data sets are not data sets, that is, they cannot in general arise from a state of the ring. If the survey is thought of as a mapping ϕ of the state space into the data space,

$$\phi: W \to \phi(W) \subset R^{2N},$$

then the dimensionality of the manifold $\phi(W)$ is smaller than 2N. In fact, it can be no larger than 2N-3, due to the fact that ϕ is not one-to-one; if W_1 and W_2 are related by an isometry, then $\phi(W_1) = \phi(W_2)$. The fibers $\{\phi^{-1}(M) \mid M \varepsilon \phi(W)\} \subset W$ partition state space into equivalence classes such that two states are in the same class provided there is an isometry connecting them. Each $\phi^{-1}(M)$ is, of course, itself a submanifold of W, but not a vector space.

Well...let's not get overheated. Global topological
(2a) properties of the manifold $\phi(W)$ undoubtedly would make pleasing objects of contemplation, but with a view to doing some calculations we must narrow our field of view to its local behavior in the neighborhood of some point, $M^* \varepsilon \phi(W)$. Of course, M^* is taken to be the survey data that would be measured if the magnets were positioned at their sites. We shall now remedy the inconvenience that $\phi(W)$ is not a vector space by constructing its tangent space at M^*. The geodetic matrix Ω is defined by the linear connection

$$M = M^* + \Omega w + \text{higher order terms in } w$$
$$\simeq M^* + \Omega w.$$

The 2Nx2N matrix Ω linearly maps state space into the tangent space at M^*, whose elements are the "differentials"

$$m = \Omega w, \qquad w \epsilon W \quad . \tag{1}$$

The definition relating m to differential forms dD_k and da_k will be given shortly.

To proceed with constructing Ω, it is easiest to take things a step at a time. First, the Wronskian $\partial(D_k a_k)/\partial(u_k v_k)$ is given by

$$\begin{array}{c} \\ D \\ a \end{array} \begin{array}{c} \begin{array}{cc} u & v \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{array} .$$

(See Fig. 3.) To calculate $\partial(D_k a_k)/\partial(u_{k\pm1} v_{k\pm1})$, use a two-step process. First construct frames $F'_{k\pm1}$ parallel to F_k. Then multiply Wronskians:

$$\frac{\partial(D_k\ a_k)}{\partial(u_{k+1}\ v_{k+1})} = \begin{array}{c} \\ D_k \\ a_k \end{array} \begin{array}{c} \begin{array}{cc} u'_{k+1} & v'_{k+1} \end{array} \\ \begin{pmatrix} -\alpha_k & 0 \\ 0 & 1 \end{pmatrix} \end{array} \cdot \begin{array}{c} \\ u'_{k+1} \\ v'_{k+1} \end{array} \begin{array}{c} \begin{array}{cc} u_{k+1} & v_{k+1} \end{array} \\ \begin{pmatrix} \cos\theta_{k+1} & -\sin\theta_{k+1} \\ \sin\theta_{k+1} & \cos\theta_{k+1} \end{pmatrix} \end{array}$$

$$= \begin{pmatrix} -\alpha_k \cos\theta_{k+1} & \alpha_k \sin\theta_{k+1} \\ \sin\theta_{k+1} & \cos\theta_{k+1} \end{pmatrix} . \tag{2}$$

10.1

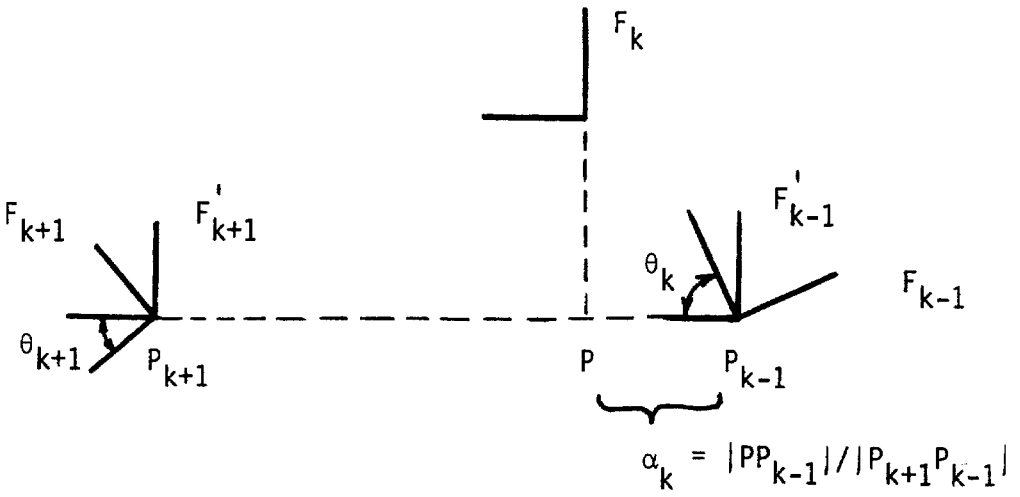

Fig. 3. Parallel translation of frames. The Wronskian $\partial(D_k\ a_k)/\partial(u_{k+1}\ v_{k+1})$ is calculated most easily in two steps. First the frame F_k is parallel translated to its neighboring sites, defining frames F'_{k+1}. These are connected to the sites' frames F_{k+1} by a rotation.

Here, α_k is the ratio:

$$\alpha_k \equiv |P_{k-1}P|/|P_{k-1}P_{k+1}| \ ,$$

where P is the intersection of the sagitta with the chord. The angle θ_{k+1}, is the angle required to rotate F_k into F_{k+1}. A similar calculation at P_{k-1}, yields the Wronskian

$$\begin{array}{c} \\ D_k \\ a_k \end{array} \begin{array}{c} \begin{array}{cc} u'_{k-1} & v'_{k-1} \end{array} \\ \begin{pmatrix} -\bar{\alpha}_k & 0 \\ 0 & -1 \end{pmatrix} \end{array} \cdot \begin{array}{c} \\ u'_{k-1} \\ v'_{k-1} \end{array} \begin{array}{c} \begin{array}{cc} u_{k-1} & v_{k-1} \end{array} \\ \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix} \end{array}$$

$$= \begin{pmatrix} -\bar{\alpha}_k\cos\theta_k & -\bar{\alpha}_k\sin\theta_k \\ \sin\theta_k & -\cos\theta_k \end{pmatrix} \ ,$$

where $\bar{\alpha}+\alpha = 1$. All other Wronskians vanish. That is,

$$\text{for } |m-k|>1, \quad \frac{\partial(D_k \ a_k)}{\partial(u_m \ v_m)} = 0.$$

The factors α and $\bar{\alpha}$ appearing in these expressions are awkward. We can get rid of at least one of them by absorption into the measurement variable. This can be done in a way that makes the Wronskian (2) unitary by replacing dD_k with $-\frac{1}{\alpha_k} dD_k$. By linearity we can next combine

all these relations into one; viz,

$$m_{;k} = \tilde{\omega}^{(k)}\, w_{;k-1} + Q^{(k)}\, w_{;k} + \omega^{(k)}\, w_{;k+1}\ ,$$

$$m_{;k} = \begin{pmatrix} -dD_k/\alpha_k \\ da_k \end{pmatrix},$$

$$\omega^{(k)} = \begin{pmatrix} \cos\theta_{k+1} & -\sin\theta_{k+1} \\ \sin\theta_{k+1} & \cos\theta_{k+1} \end{pmatrix}, \qquad (3)$$

$$\tilde{\omega}^{(k)} = \begin{pmatrix} (\bar{\alpha}_k/\alpha_k)\cos\theta_k & (\bar{\alpha}_k/\alpha_k)\sin\theta_k \\ \sin\theta_k & -\cos\theta_k \end{pmatrix},$$

$$Q^{(k)} = \begin{pmatrix} -1/\alpha_k & 0 \\ 0 & 0 \end{pmatrix}.$$

The matrix ω is orthogonal, but $\tilde{\omega}$ is not, except for the special case $\bar{\alpha} = \alpha = 1/2$.

Provided that we accept the convention of site indices obeying mod N arithmetic, the recursion in Eq.'s (3) is satisfied for all k in the ring. By collecting them together, we obtain the full matrix Ω written in block form,

$$\Omega = \begin{bmatrix} Q^{(1)} & \omega^{(1)} & 0 & 0 & \cdots & 0 & 0 & \tilde{\omega}^{(1)} \\ \tilde{\omega}^{(2)} & Q^{(2)} & \omega^{(2)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \tilde{\omega}^{(3)} & Q^{(3)} & \omega^{(3)} & \cdots & 0 & 0 & 0 \\ \vdots & & & & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & \tilde{\omega}^{(N-1)} & Q^{(N-1)} & \omega^{(N-1)} \\ \omega^{(N)} & 0 & 0 & 0 & \cdots & 0 & \tilde{\omega}^{(N)} & Q^{(N)} \end{bmatrix}$$

or in block components,

$$\Omega_{;ik} = \tilde{\omega}^{(i)}\ \delta_{i\ k+1} + Q^{(i)}\ \delta_{ik} + \omega^{(i)}\ \delta_{i\ k-1}. \qquad (3')$$

If the sites are designed to be equally spaced around a circular ring, the problem becomes formally easier. We will label this configuration C_N, after the cyclic group of order N. For C_N the superscripts on the 2x2 sub-matrices of Ω can be dropped, since

$$\forall k: \quad \omega^{(k)} = \omega \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

$$\tilde{\omega}^{(k)} = \tilde{\omega} \equiv \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}, \qquad (4)$$

$$Q^{(k)} = Q \equiv \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that we also have $\alpha_k = \bar{\alpha}_k = 1/2$, for all k, so that

$$m_{;k} = \begin{pmatrix} -2\ dD_k \\ da_k \end{pmatrix}.$$

Both ω and $\tilde{\omega}$ are now orthogonal. In fact,

$$\tilde{\omega} = \sigma_3 \omega^T = \omega \sigma_3,$$

where σ_3 is the usual Pauli matrix. Further, each is a .

N^{th} root of unity

$$\omega^N = \tilde{\omega}^N = 1.$$

For the C_N ring, each (block) row of Ω is obtained from its predecessor by cyclically shifting the (block) elements by one (block) column; that is,

$$\forall\ i,k,m:\quad \Omega_{;i+m\ k+m} = \Omega_{;ik}\quad .$$

A matrix possessing such structure is called "block circulant".

The state w could be obtained directly by multiplying Eq. (1) on the left by Ω^{-1} were it not for the awkward circumstance that the geodetic matrix is singular. This is, of course, due to the fact that the survey data are invariant under isometries. Under linearization this means that there are vectors $s \varepsilon W$ that are annihilated by Ω:

$$\Omega s = 0.$$

As usual, the set of all such vectors is itself a (vector subspace of W called the "null space" of Ω, which we shall symbolize as $\underline{\text{null}}(\Omega)$; its orthogonal subspace will be called the "domain" of Ω, symbolized $D(\Omega)$.[1] These are defined also by the conditions:

[1] If Ω is viewed algebraically, as a homomorphism, then $\underline{\text{null}}(\Omega)$ is its kernel.

$$s \in \underline{\text{null}}(\Omega) \rightarrow \Omega s = 0;$$

$$\underline{\text{null}}(\Omega) \cap D(\Omega) = \emptyset;$$

$$\underline{\text{null}}(\Omega) \cup D(\Omega) = W.$$

Dimensionality of $\underline{\text{null}}(\Omega)$ must be at least three, since three linearly independent vectors can be generated by an infinitesimal rotation of the lattice about some central point and translations along two orthogonal directions. These isometries are true global symmetries of the problem, independent of the ring's configuration. In addition, the C_N ring possesses an <u>accidental</u> local symmetry when N is even, raising the dimensionality to four. (This will be demonstrated shortly.) The conjectures that this is the maximum value and that it occurs only for C_N with N even are supported by numerical experimentation, but they have not been proved.

Let us demonstrate explicitly the null vectors of Ω in C_N. Using the block format described previously, we have

$$\begin{aligned} \forall m: \quad (\Omega s)_{;m} &= \Omega_{;mk}\, s_{;k} \\ &= (\tilde{\omega}\, \delta_{m\, k+1} + Q\, \delta_{mk} + \omega\, \delta_{m\, k-1}) s_{;k} \\ &= \tilde{\omega}\, s_{;m-1} + Q\, s_{;m} + \omega\, s_{;m+1} \\ &\equiv 0. \end{aligned} \tag{5}$$

The simplest symmetry is (infinitesimal) rotation of the lattice about its center. This amounts to setting

$$\forall\, k: \quad du_k = 0 \quad , \quad dv_k = 1,$$

that is,

$$\forall k: \quad s_{;k} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

Using Eq.'s (4), this gives us

$$(\Omega s)_{;m} = \begin{pmatrix} \sin\theta \\ -\cos\theta \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} ,$$

proving that it is indeed a null vector.

Translations are a little more involved. The state s represents a translation provided that

$$\forall\, k: \quad s_{;k+1} = \omega^T s_{;k} \ .$$

By iterating, we get easily

$$s_{;k} = (\omega^T)^{k-1} s_{;1} \ .$$

(Since $(\omega^T)^N = 1$, the mod N subscript convention is satisfied.)

Then, using $\tilde{\omega} = \sigma_3 \omega^T$,

$$(\Omega s)_{;m} = \left[\sigma_3 \omega^T (\omega^T)^{m-2} + Q(\omega^T)^{m-1} + \omega(\omega^T)^m \right] s_{;1}$$

$$= \left[\sigma_3 \, \omega^T \omega + Q + \omega \omega^T \right] (\omega^T)^{m-1} \, s_{;1}$$

$$= \left[\sigma_3 + Q + 1 \right] (\omega^T)^{m-1} \, s_{;1}$$

$$= 0.$$

There are two linearly independent choices for $s_{;1}$, and correspondingly the translation part of $\underline{\text{null}}(\Omega)$ is two-dimensional.

Interestingly enough, there is yet another symmetry when N is even. This one can be associated with exciting the N/2 harmonic variation in azimuth and is obtained by setting

$$s_{;k} = \begin{pmatrix} 0 \\ (-1)^k \end{pmatrix}.$$

Putting this into Eq.(5),

$$(\Omega s)_{;m} = (-1)^m \left(\tilde{\omega} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + Q \begin{pmatrix} 0 \\ -1 \end{pmatrix} + \omega \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = 0.$$

Unlike translations and rotation, this "skew" symmetry is <u>accidental</u>: it arises from the particular configuration

and is valid to first order only. It is not an isometry; the qualification made earlier that the primary survey variables are "almost always" sufficient to specify a state is due to its existence. The measurement of a single nearest-neighbor chord would be enough to break this symmetry.

For the general ring, null(Ω) is still computable provided that it is spanned only by isometries: one simply constructs state vectors representing infinitesimal translations and rotations.

Fourier analysis provides a useful, familiar tool for processing data on a C_N ring. Its power in this context is derived from the fact that the unitary Fourier transformation block-diagonalizes Ω. Define the 2Nx2N transform matrix

$$S_{;km} \equiv \sqrt{1/N}\, \mathbb{1}\, \xi^{km},$$

where $\xi = \exp(2\pi i/N)$, and $\mathbb{1}$ is the 2x2 unit matrix.[2] It is easy to verify, using Eq.(3'), that

$$(S^{+}\Omega S)_{;km} = \delta_{km}\, 2W(k),$$

$$W(k) = \frac{1}{2}\,(\tilde{\omega}\,\xi^{-k} + Q + \omega\,\xi^{k}) \qquad (6)$$

$$= \begin{bmatrix} \cos\theta\,\cos k\theta - 1 & -i\,\sin\theta\,\sin k\theta \\ \sin\theta\,\cos k\theta & i\,\cos\theta\,\sin k\theta \end{bmatrix}$$

[2] Actually, any matrix would do.

with $\theta = 2\pi/N$. The important point here is the δ_{km} factor: by transforming the problem, a complicated system of equations of order 2N has been split into N simpler subsystems of order 2. Each of these is solved easily by inversion, provided that

$$\det(W) = i \sin k\theta \, [\cos k\theta - \cos\theta]$$

is not zero. Singularities (det W = 0) occur for k = 0, k = 1, and k = N/2. These obviously are associated with the rotation, translation, and skew null vectors of the preceding discussion.

By carrying out the mechanics of confirming Eq.(6) it becomes clear that S will block-diagonalize any block circulant matrix. Thus, the usefulness of Fourier analysis in this problem is fundamentally a consequence of its C_N symmetry: in general, taking a Fourier transform will not block-diagonalize Ω.

In order to extend the method to general configurations it seems natural to try replacing the Fourier set of basis vectors with the true eigenvectors of Ω. There is an objection to this, however: because Ω is not symmetric, its eigenvectors need not be real or orthogonal to one another. This is not catastrophic; algebraically, there is nothing wrong with a complex non-orthonormal basis; numerically, it is awkward and may lead to accuracy problems. Making a singular value decomposition of Ω is a way to get around this.

It is interesting that, with all the attention paid to eigenvalues and eigenvectors in the academic physics curriculum, singular value decomposition is ignored. This is probably a consequence of the emphasis placed on diagonalizing Hermitian matrices, representing observable operators. It seems natural to ask whether one can diagonalize a non-Hermitian matrix, yet the question is generally not addressed.

The complete statement of the theorem defining singular value decomposition is this: Given any mxn matrix A with $m \geq n$ there exist matrices U(mxn), Σ(nxn) and V(nxn) such that

$$A = U\Sigma V^{+}, \quad (V^{+} = \text{Hermitian conjugate of } V) \tag{7}$$
$$U^{+}U = V^{+}V = VV^{+} = \mathbf{1}_{n},$$
$$\Sigma = \underline{\text{diag}}\ (\sigma_1, \sigma_2, \ldots \sigma_n).$$

The matrix V is unitary; U is "generalized unitary" in the sense that it is not necessarily square. The non-zero, diagonal elements of Σ are the singular values of A; they are obviously square roots of the eigenvalues of $A^{+}A$; it is conventional to use the positive square roots and to order them so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. Of course, if $m \leq n$ one has an obvious corollary by taking the Hermitian conjugate (2b) of Eq.(7).

The columns of U and V comprise a set of modes or orthonormal basis vectors in the tangent plane and state

space, respectively, that is the generalization appropriate to the given ring of the harmonic basis of Fourier analysis. Each mode in state space excites one and only one mode in the tangent space;

$$\Omega = U\Omega_D V^+ \rightarrow \Omega_D(V^+w) = U^+m \quad ,$$

where Ω_D is diagonal.

In addition to orthogonality, this representation of Ω possesses the advantage that the singular values are real; (3) in general, the eigenvalues of Ω would be complex. That different bases are required in the two spaces is the price one pays for dallying with a non-Hermitian matrix.

Although useful, singular value decomposition is really peripheral to the main theme, which is to "solve" Eq.(1).

Section 3 The Moore-Penrose pseudoinverse

Singularity of Ω means that the equation

$$\Omega w = m$$

has either no solution or a manifold of solutions, depending on whether m is in the range of Ω. In the latter case, the manifold arise from the fact that if w* is a solution, then so is w*+w', where w' is any vector in the null space of Ω. We seek an unambiguous linear map $X:R^{2N} \rightarrow W$ such that $w = Xm$ will, in some sense, be our accepted solution

of the state for arbitrary m. One obvious criterion is that the measurement defect, $\Omega X \cdot m - m$, should be small. If we use a L^2 norm, then this becomes:

$$\text{choose } X \text{ to minimize } ||(\Omega X - 1)m||$$

which implies that X must satisfy

$$\Omega X = P_{R(\Omega)}, \quad \text{or} \quad \Omega^T \Omega X = \Omega^T \tag{8}$$

where $P_{R(\Omega)}$ is the projection operator onto the range of Ω.

Interestingly enough, this is not sufficient to determine X uniquely, unlike most "least squares" problems. For example in doing a least squares fit of a theoretical expression to experimental data, one is imbedding a fairly low dimensional manifold - generated by allowing the parameters of the theory to vary over their allowed range - into a much higher dimensional one - the space of all possible data. Unless the surface intersects itself, the imbedding is usually one-to-one: each point on the theory's surface corresponds to exactly one set of parameter values. Thus, if there is a unique point on the surface closest to measured data, then there will be a unique associated assignment of parameters. The situation is fundamentally different here; the imbedding does not preserve

dimensionality. Finding the solution in the measurement space specifies not a state but a manifold of states.

So, we need another condition that selects a unique state from this manifold. The natural one is to demand that the preferred state be orthogonal to null(Ω). Clearly, this is appealing on heuristic grounds: the estimated state should not contain any components that do not alter the survey data from their design values. This is also equivalent to requiring it to have minimum norm, since the manifold of solutions is parallel to null(Ω). That is appealing too; it means that if one wanted to zero the state by repositioning magnets, then the total movement of magnets would be kept to a minimum, in the sense of a L_2 norm. This condition can be expressed algebraically by the equation

$$P_{D(\Omega)}X = X, \tag{9}$$

where $P_{D(\Omega)}$ is the projector onto $D(\Omega)$.

For any matrix Ω there is one and only one solution for X in Eq.'s (8) and (9): the Moore-Penrose pseudoinverse, (4) symbolized $\Omega^{\div}$. It reduces to the inverse Ω^{-1} when Ω is non-singular, and if Ω is rectangular with $\Omega^T\Omega$ non-singular, then $\Omega^{\div}$ is the usual least-squares matrix $(\Omega^T\Omega)^{-1}\,\Omega^T$. In our particular application neither of these conditions holds, and the problem of computing $\Omega^{\div}$ is more involved.

Standard programs now exist, but they are somewhat inefficient when Ω is sparse, as is the case here. This is especially true for the C_N ring, for which Ω is block circulant. The pseudoinverse is then block circulant also, considerably reducing the complexity of its computation. We will not treat the computational aspects in detail here; their denouement is postponed to another memo. For now it suffices that

$$X = \Omega^{\div}$$

is the unique solution to Eq.'s (8) and (9).

When Ω is the direct sum of an invertible matrix, Λ, and a zero matrix,

$$\Omega = \left(\begin{array}{c|c} \Lambda & 0 \\ \hline 0 & 0 \end{array}\right) ,$$

computing $\Omega^{\div}$ becomes a trivial matter:

$$\Omega^{\div} = \left(\begin{array}{c|c} \Lambda^{-1} & 0 \\ \hline 0 & 0 \end{array}\right) .$$

(It is easily verified that $\Omega^{\div}$ obeys the two defining Eq.'s (8) and (9).) In particular, if $\Lambda = \underline{\text{diag}}\ (\lambda_1\lambda_2\ldots\lambda_r)$, then $\Omega^{\div} = \underline{\text{diag}}\ (\lambda_1^{-1}\lambda_2^{-1}\ldots\lambda_r^{-1}00\ldots0)$. This is useful in conjunction with the singular value decomposition of Ω; it enables us to calculate $\Omega^{\div}$ via

$$\Omega = V\Omega_D U^{+} \rightarrow \Omega^{\div} = U\Omega_D^{\div} V^{+} . \qquad (10)$$

However, if one only wants to calculate $\Omega^{\div}$ and doesn't care about the basis vectors per se, this is not an efficient way to go about it. It would be comparable to calculating the eigenvalues and eigenvectors of a matrix in order to invert it; the former requires an (truncated) infinite procedure, while the latter is finite. Eq.(10) is useful, nonetheless, for selectively filtering or purging unwanted "generalized harmonics" of the data and solutions.

It was known previously that measurement errors in survey data could lead to substantial low harmonic components in magnet positions. If desired, these could be artificially suppressed by deleting the low harmonic content of the data, or by weighting the Fourier components in such a way that the higher harmonics are reproduced exactly but the lower are suppressed. Singular value decomposition leads to a natural extension of this for the general lattice. Using Eq.(10) we write

$$w = \Omega^{\div} m = Um' \quad , \quad \text{where } m' = \Omega_D^{\div} V^{+} m \quad . \qquad (11)$$

The state w is a superposition of the columns of U weighted according to the elements of m'. These columns - or equivalently, the eigenvectors of $\Omega^{+}\Omega$ - form a natural basis for state space, and they reduce to the harmonic basis when the ring is C_N. We can associate with each

one its "sequency", an appropriate generalization of the concept of frequency, which is the number of times neighboring elements have opposite signs, i.e., the number of times that $U_{.k}$ changes sign. Suppressing the low sequency singular vectors is then the analog to suppressing low frequency components of a harmonic expansion. One simply zeroes or weights appropriate elements of m' just as before. The only real difference is that the basis states are different in measurement space and state space, since in general $V \neq U$.

Section 4 Closed orbits

Once the state has been estimated a decision must be made on whether to actually move magnets to new positions. The subject of errors plays an important role in this. Originally, the magnets are located to within some placement error σ_o of their design sites. It seems reasonable to assume that the original placement error of each magnet is isotropically distributed about its site and that there are no correlations between sites. Survey data are then taken, and this introduces another source of error: the measurements themselves. Because of these the state estimate will be different from zero even when the magnets have been positioned perfectly. Using the usual ansatz that models the errors as an additive random variable, we can write the data vector m as

$$m = \langle m \rangle + \hat{m},$$